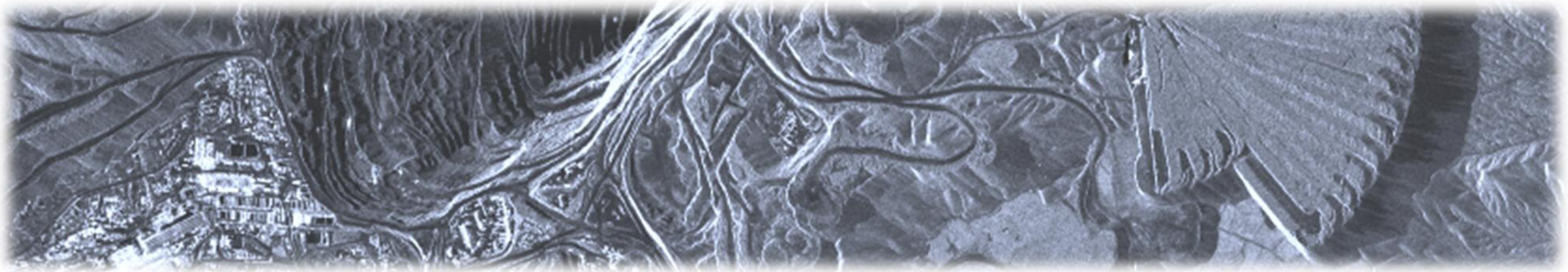


Complex statistics of Synthetic Aperture Radar data using high resolution TerraSAR-X images

Anca Popescu, Mihai Datcu, Madalin Frunzete

The 5th TerraSAR-X Science Team Meeting

DLR, 2013

CEOSpaceTech

Outline

1. Problem statement – SAR data modeling
2. The increased resolution problem
3. General assumptions on SAR data statistics
4. Statistical testing
    1. Experimental data
    2. Results and discussion

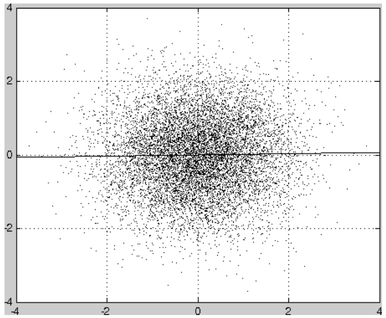
7. Conclusions

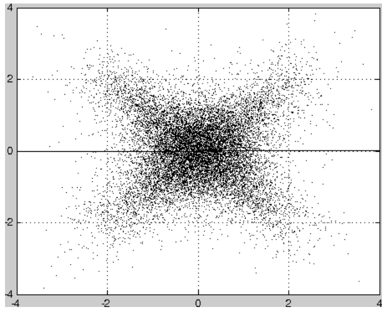
1. Problem statement – SAR data modeling

Generally the complex SAR signal is treated as proper circular random process, under the assumption that the real and imaginary parts are independent Gaussian distributed random variables.

However, in practice, this assumption rarely validates, especially with very high resolution data.

The implications are on the usage of special statistics adapted for complex signals which do not follow the circularity assumptions.





CE SpaceTech

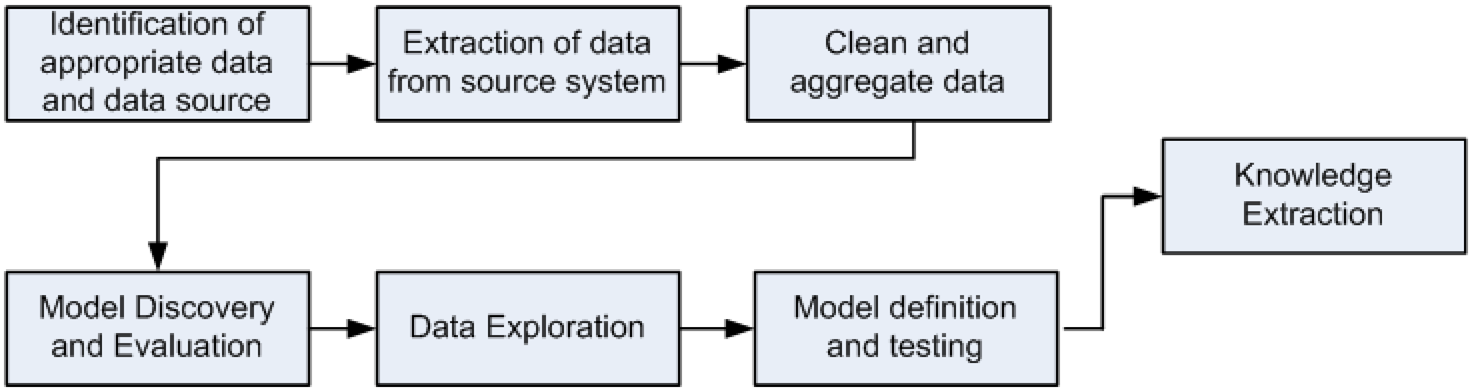
1. Problem statement – SAR data modeling

Information in EO SAR data:
- Geometry
- Patterns and structures
- Time changes
- Texture
- Artifacts and distorssions
- Phase ….

✓Most of the SAR processing tools rely on assumed data models and preprocess data
✓Most of the remote sensing applications make use of data statistics
✓Huge amount of data used in KDD, which depend on correctly assessed data properties

1. Problem statement – SAR data modeling

KDD Process



Model definition is essential for Knowledge Extraction – training, fitting, tuning - prediction error

CEOSpaceTech

1. Problem statement – SAR data modeling

Increased number of applications and techniques, increased number of missions, higher resolution, higher quality, increased information
Smaller observation time intervals for the signal to be assumed stationary.

$$f(\zeta) \longrightarrow f(x,\theta \mid \zeta) \longrightarrow f(x,\theta) \longrightarrow y = H(x,\theta) + n \longrightarrow f(y \mid x,\theta)$$

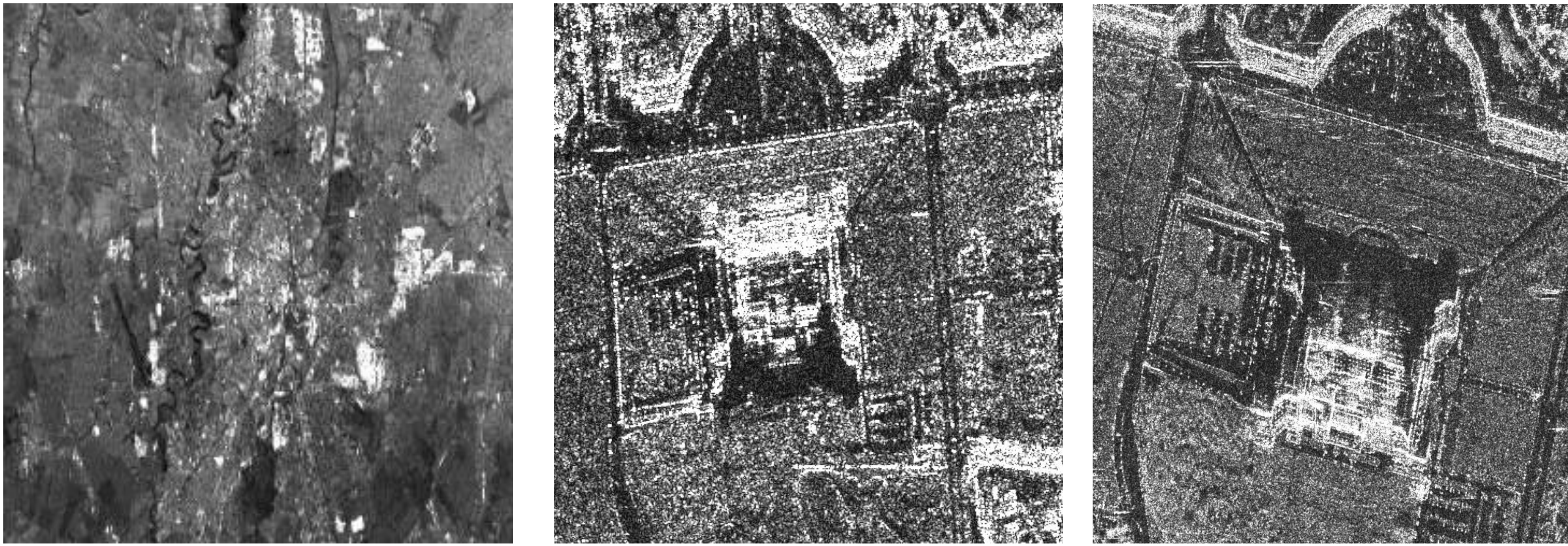
Scene | Cross-section | Sensor - H | Image

$\longleftarrow$ Scene Information $\longrightarrow$

$\longleftarrow$ Image Information $\longrightarrow$

CEOSpaceTech

2. The increased resolution problem

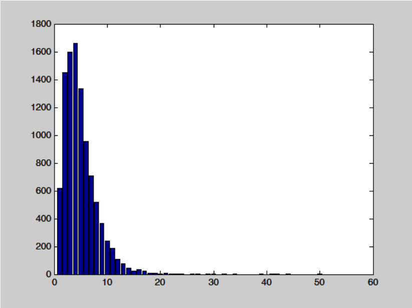


ERS – 1 , TerraSAR-X StripMap and TerraSAR-X High Resolution Spotlight of Bucharest and House of Parliament

3. General assumptions on SAR data statistics

- The speckle intensity after multi-looking is generally modeled as a Gamma distributed random variable – assumption holds for homogeneous areas

$$f_{Yi}(y) = \frac{n^n}{\Gamma(n)} y^{n-1} \exp(-ny), \quad n, y > 0$$



- Multiplicative model for speckle to take into account the image formation process (valid for all coherent data acquisitions)

- The bivariate Gaussian model assumes real and imaginary parts independent and identically distributed zero mean random variables (1 look – Rayleigh distribution, n-looks – Gamma distribution)

3. General assumptions on SAR data statistics

- The Gaussian distribution is the most commonly employed model for the observed data, mainly to facilitate computation – seldom confirmed in practice

- Intensity backscatter is modeled as generalized inverse Gaussian, particularized for different scattering cases (homogeneous – Gamma, heterogeneous – K-distribution, very heterogeneous – G-Zero distribution)

- Complex statistics – pseudo-covariance, pseudo – power spectral density, almost never employed

$$Cov_z[k_1,k_2]=E\{z[k_1]z^*[k_2]\}$$
$$\tilde{C}ov_z[k_1,k_2]=E\{z[k_1]z[k_2]\}$$

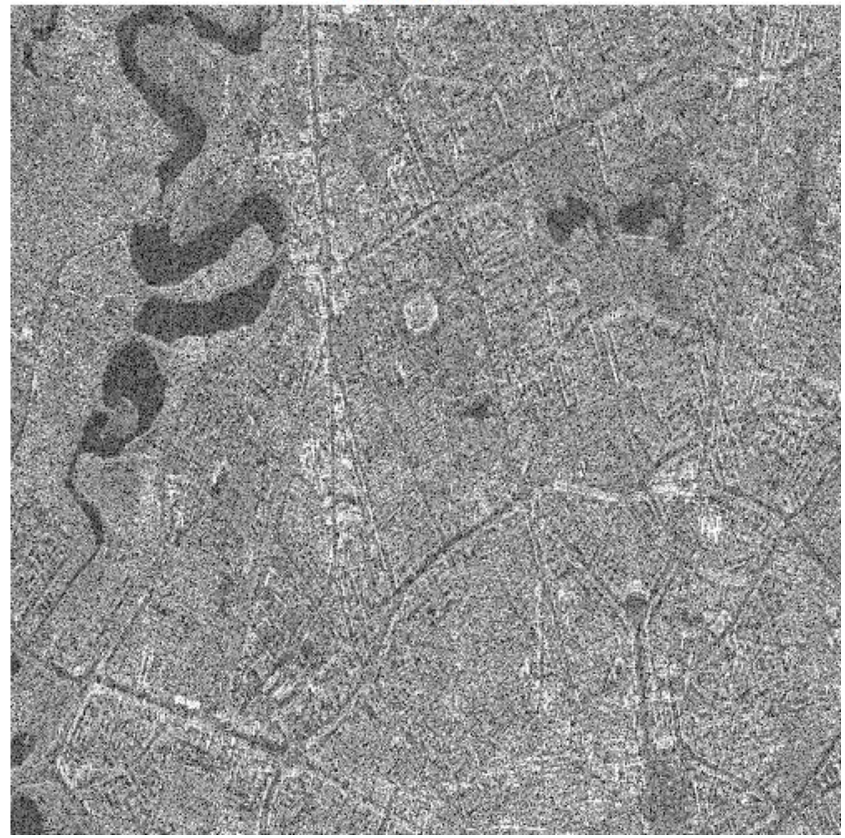
4. Statistical testing

1. Distribution fitting on a class-dependent approach
2. Zero Mean Gaussian assumption for real and imaginary parts
3. Goodness of fit for Gaussian distribution test
4. Statistical independence test
5. Pseudo-covariance computation and assessment

4. Statistical testing – Data characteristics

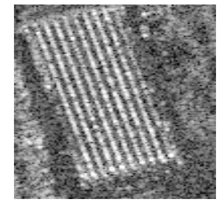
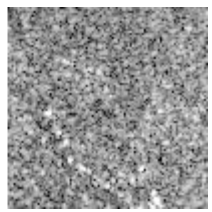
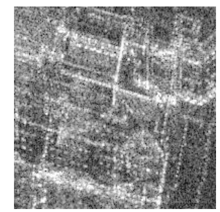
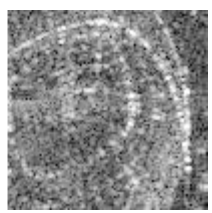
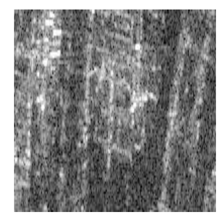
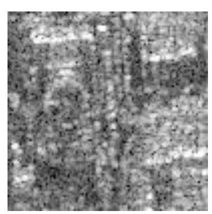
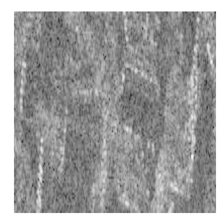
SpotLight preview

StripMap preview



CEOSpaceTech

4. Statistical testing – Data characteristics

Tested classes Preview			
Water		Industrial Area	
Park		Parliament	
Sports Arena		Apartment blocks	
Road		Houses	

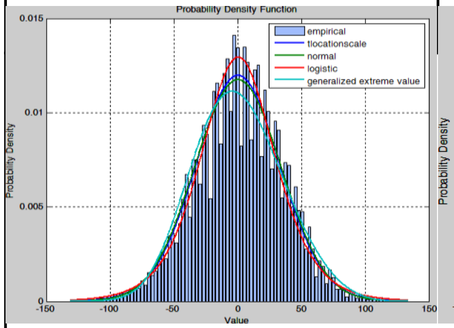
4. Statistical testing – Distribution fitting on a class-dependent approach

- Fit all valid distributions to the data (for real and imaginary parts, amplitude and intensity)
- Estimation performed based on Bayes Information Criterion (BIC) and Akaike Information Criterion (AIC)
- Evaluate distribution parameters
- Fitted distributions : Normal, t-location scale, Logistic, Exponential, Generalized Extreme Value, Generalized pareto

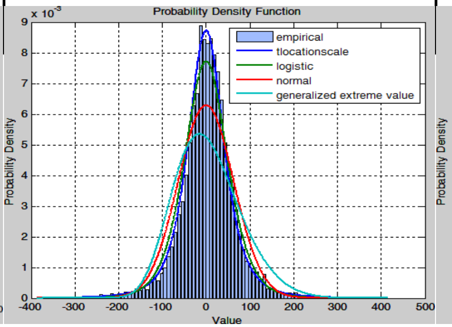
4. Statistical testing – Distribution Fitting

StripMap

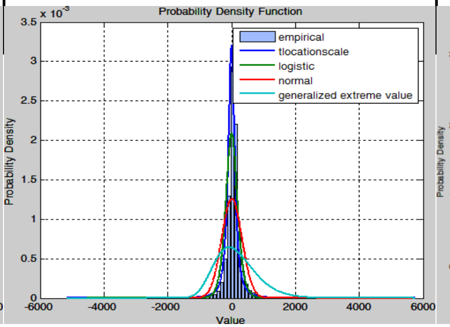
Water



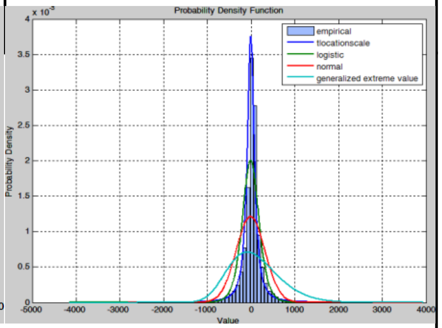
Park



Sports Arena



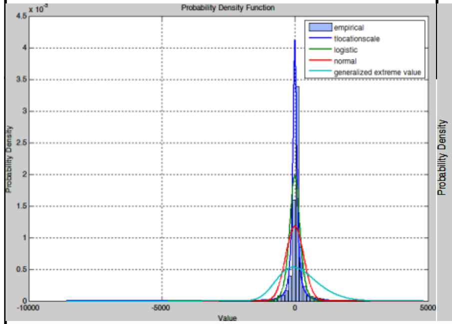
Road



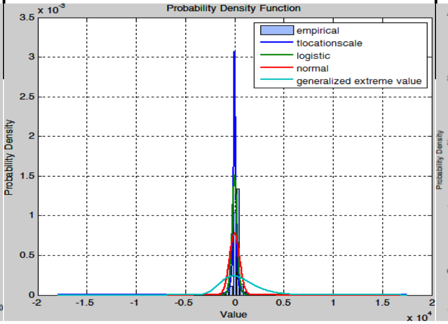
StripMap

Spotlight

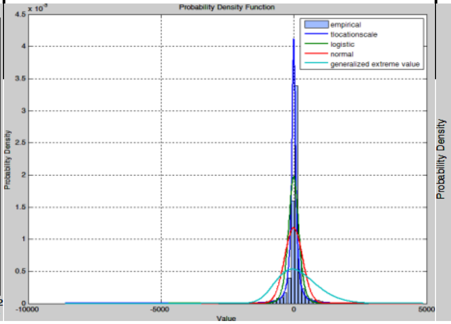
Industrial Area



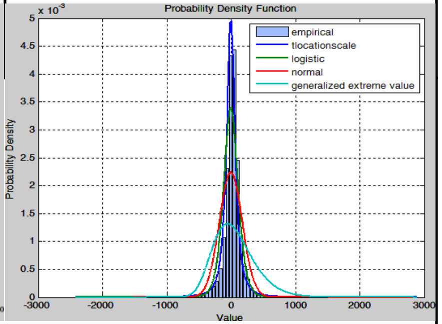
Parliament



Apartment



Houses



Spotlight

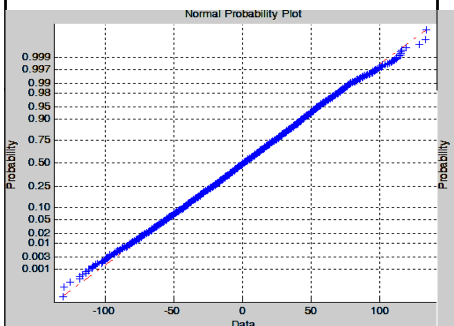
CEOSpaceTech

4. Statistical testing – Hypothesis testing (normal distribution)

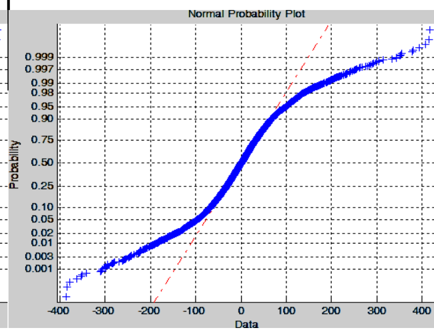
Test the real and imaginary parts for SM and HS samples from 8 different classes against the hypothesis that they come from a normal distribution.

StripMap

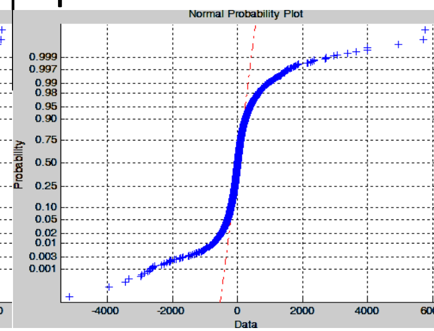
Water



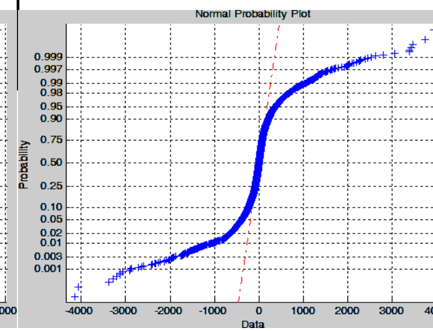
Park



Sports Arena

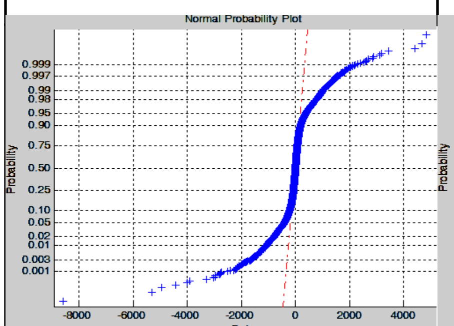


Road

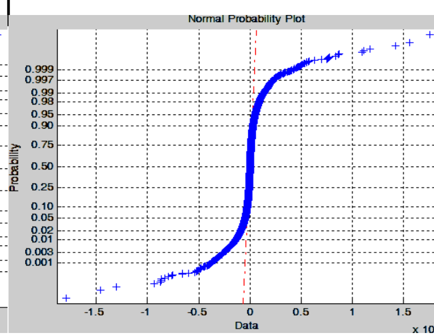


Spotlight

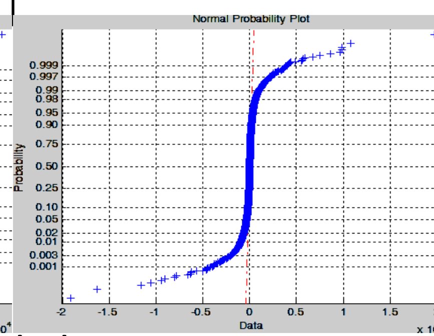
Industrial Area



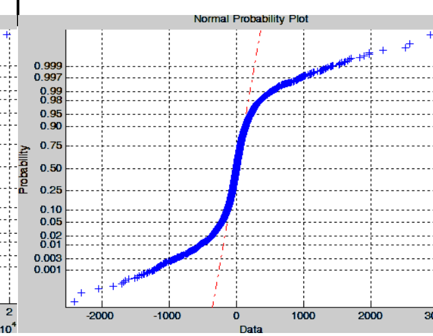
Parliament



Apartment blocks



Houses



CEOSpaceTech

4. Statistical testing – Hypothesis testing (Gaussian distribution)

Chi-square goodness-of-fit test the real and imaginary parts for SM and HS samples from 8 different classes against the hypothesis that they come from a normal distribution.

Tested classes Chi-square test score (null hypothesis rejected at 5 % significance level)					
SM	Chi-square Test Score (1 = fail)	T-Test Score (zero-mean normal distribution)	HS	Chi-square Test Score (1 = fail)	T-Test Score (zero-mean normal distribution)
Water	1	0	Water	1	0
Park	1	0	Park	1	1
Grassland	1	0	Industrial Area	1	1
Sports Arena	1	1	Parliament	1	0
Road	1	1	Apartment blocks	1	0
Apartment blocks	1	0	Houses	1	0

4. Statistical testing – Statistical Independence

Test the independence of real and imaginary parts.
Hypotheses:

H0: real (R) and imaginary (I) parts are statistically independent
H1: real and imaginary parts are dependent

Two successive transforms are applied on each of the two investigated random variables : R' and I' are are uniformly distributed in the (0; 1) interval, and U and V are standard normal random variables. By this approach, the entire independence analysis is shifted by the two transforms from the (R, I) coordinates to (R', I') coordinates and finally to the (u, v) coordinates.

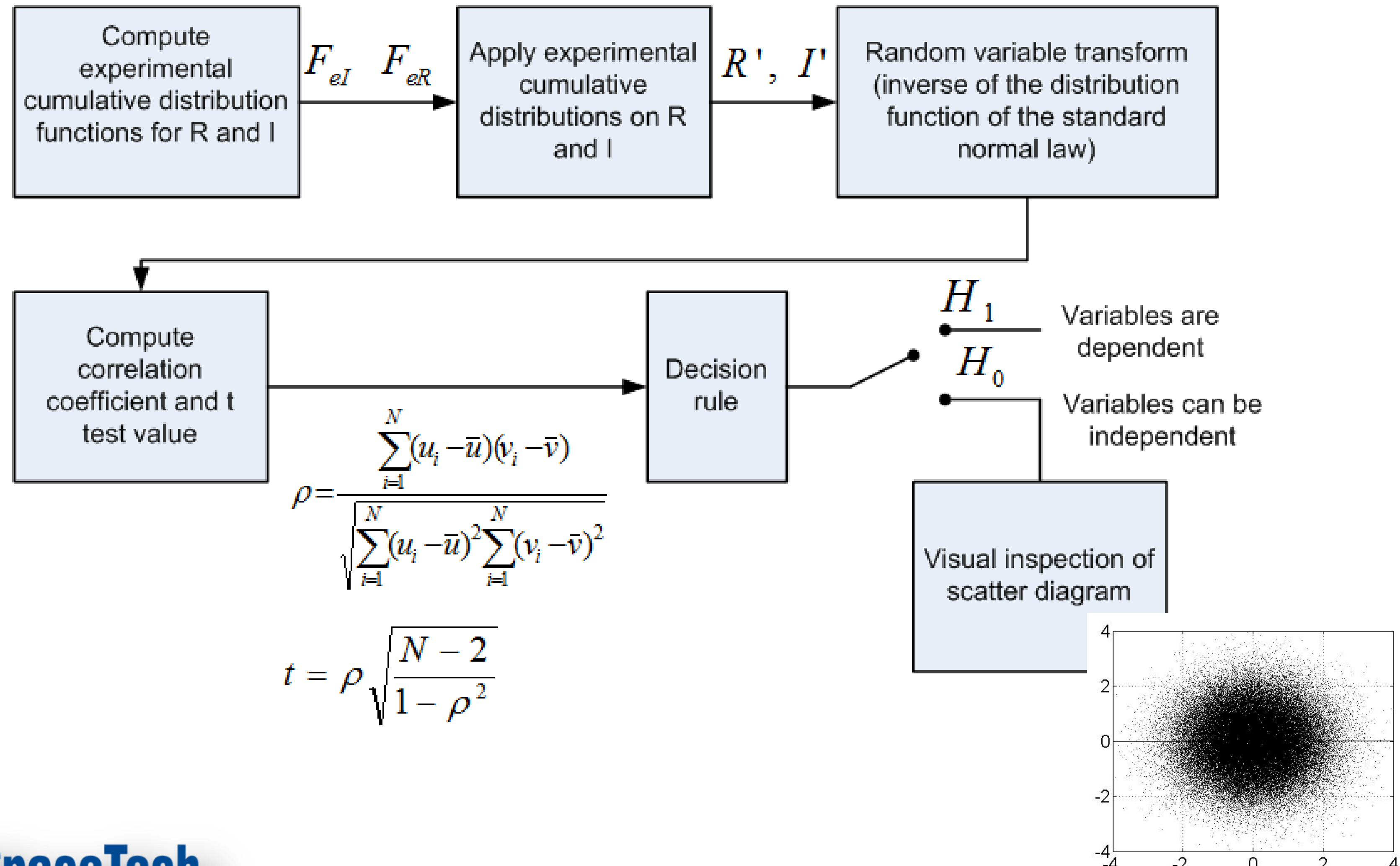
$$R \to R' \to U$$

$$I \to I' \to V$$

CEOSpaceTech

4. Statistical testing – Statistical Independence

Algorithm:



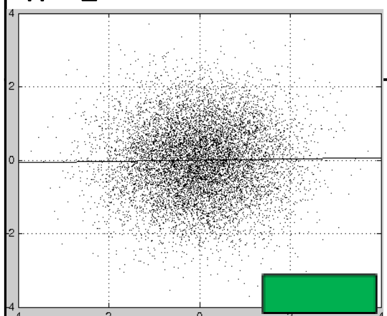
$$\rho = \frac{\sum_{i=1}^{N}(u_i - \bar{u})(v_i - \bar{v})}{\sqrt{\sum_{i=1}^{N}(u_i - \bar{u})^2 \sum_{i=1}^{N}(v_i - \bar{v})^2}}$$

$$t = \rho \sqrt{\frac{N-2}{1-\rho^2}}$$

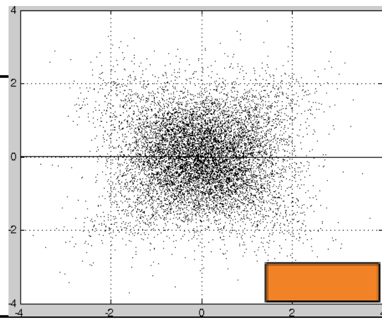
CEOSpaceTech

4. Statistical testing – Statistical Independence

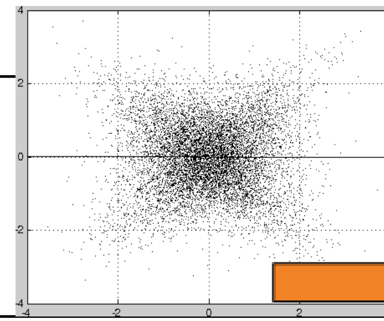
Water
Corr = 0,0173
t = 1,7313
h= 1



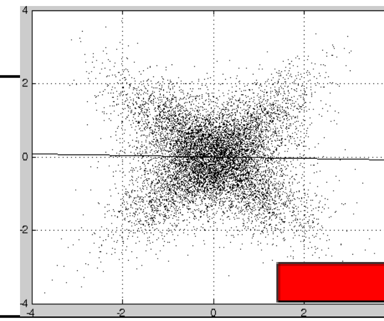
Park
Corr = -0,0041
t = 0,4199
h= 1



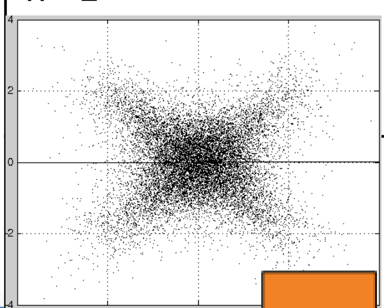
Sports Arena
Corr = -0.0026
t = 0.2585
h = 1



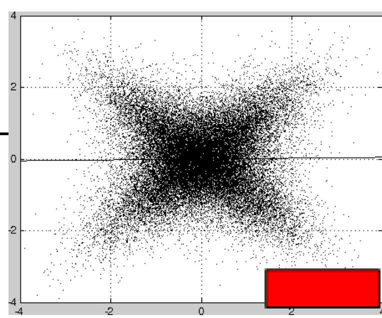
Road
Corr = -0.0205
t = 2,0512
h = 0



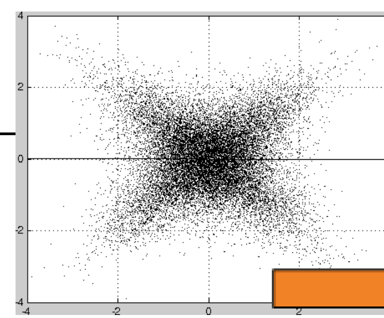
Industrial Area
Corr = 0,0013
t = 0,1611
h= 1



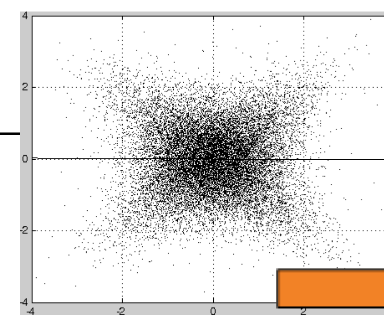
Parliament
Corr = 0,122
t =2,4339
h= 0



Apartment blocks
Corr = -0,0012
t =0,1733
h= 1



Houses
Corr = -0,0061
t =0,8565
h= 1



PASS FAIL Visual FAIL

CEOSpaceTech

4. Complex statistics – pseudo covariance

Second order statistics of complex signals are usually described by the covariance function :

$$z(t) = [x(t), y(t)]^T, \qquad z(t) = x(t) + i \cdot y(t)$$

$$Cov_z(t_1, t_2) = E\{z(t_1) z^*(t_2)\}$$

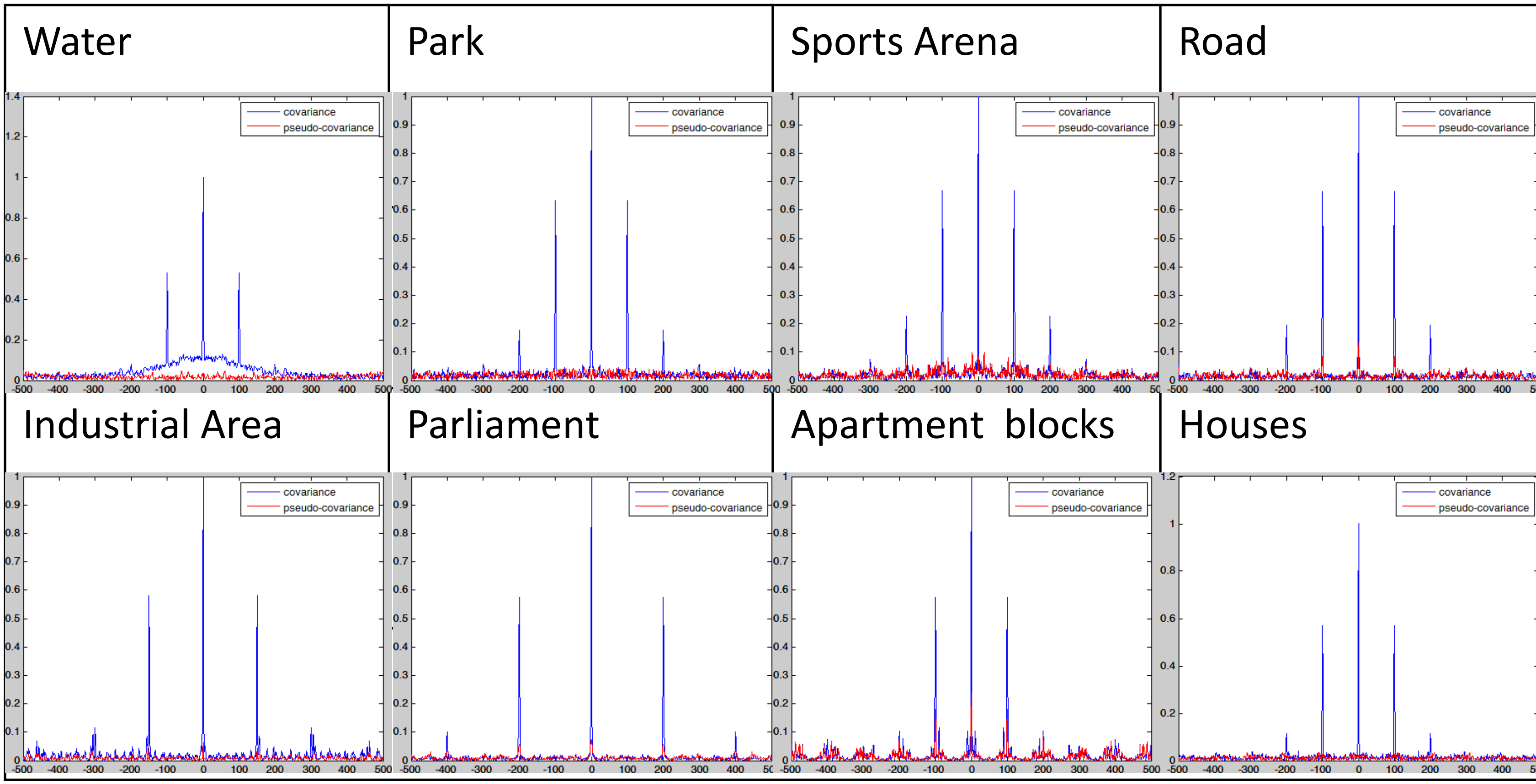
The covariance function is not always sufficient to completely describe the second order statistics. The pseudo-covariance function is defined as:

$$\widetilde{C}ov_z(t_1, t_2) = E\{z(t_1) z(t_2)\}$$

If the signal is circular, then the pseudo-covariance function vanishes and the covariance is sufficient. **A second-order circular signal is a signal whose second-order statistics are invariant in any phase transformation**.

4. Complex statistics – pseudo covariance

Circularity test for SM and HS data:



4. Conclusions

- Even with reduced analysis window size, the increased heterogeneity of high resolution SAR image signal leads to failure in most hypothesis testing scenarios.
- Areas which appear to be homogeneous and ensure WSS fail statistical tests in most of the cases
- Heterogeneous and man-made structures never obey generally accepted statistical hypotheses
- In order to describe the second-order properties of *complex* random signals completely, it is necessary to use two moments: the classical covariance function, and the pseudo-covariance function. As a direct consequence, PSD and pseudo-PSD should be employed
- WSS should be extended to SOS (second order stationarity), to take into account the pseudo-covariance